

ABSTRACT

A method, operating model, system, data structure, computer program and computer program product for analyzing and categorizing unstructured information is provided such that conventional structured data access techniques can be utilized over unstructured objects. A analysis and categorization engine builds a set of concept groupings, each grouping consisting of related words and phrases. The concept groupings are augmented by user input. A set of categories is built. The analysis and categorization engine generates a vector representation of each object based on concepts and utilizes a statistical analysis to select concepts that represent each object and assign objects to categories. Information about users, objects, and categories is stored in an open architecture, such as a relational database. An object concept based search is provided to efficiently locate meaningful objects and to provide for updating of the object categorization based on search entries.